4 Strategic Experimentation: the Undiscounted Case

PATRICK BOLTON AND CHRISTOPHER HARRIS

1. Introduction

Bolton and Harris (1999) consider a game of strategic experimentation. In this game, each player divides her time in any given period between a 'safe' action and a 'risky' action. The underlying pay-off of the safe action is known. The underlying pay-off of the risky action is unknown, and can be higher or lower than that of the safe action. The actual pay-off received by any player from an action is the underlying pay-off plus noise. Once players' actions have been chosen and pay-offs realized all players observe all choices and pay-offs. Through these observations they are able to learn about the underlying pay-off of the risky action and thus revise their common beliefs and optimal choice of action. In other words, players can learn from others' current experimentation as well as their own.

Bolton and Harris provide a complete characterization of the team solution of the strategic-experimentation game, establish the existence of a unique symmetric stationary Markov-perfect equilibrium, and show how equilibrium experimentation and pay-offs vary with the number of players and the discount rate. In particular, they show that there are two major effects at work in their model, a free-rider effect and an encouragement effect. The free-rider effect arises because experimentation is a public good. The encouragement effect arises because the prospect of future experimentation by others gives a player an incentive to increase her current experimentation in order to bring forward the time at which the additional information obtained from the others' experimentation becomes available. However, Bolton and Harris only obtain a partial characterization of the unique symmetric equilibrium, and do not provide any characterization of the set of asymmetric equilibria.

In this paper, we pursue further the analysis of strategic experimentation initiated in Bolton and Harris. We begin by making an important change in the model: we add background information in the form of an exogenous noisy signal of the underlying pay-off of the risky action. This ensures that players' preferences have a well-defined limit as the discount rate converges to zero.

The undiscounted case is much easier to analyse than the discounted case. We are therefore able to give a detailed characterization of all the equilibria of the model, both symmetric and asymmetric. Using this characterization, we can show that the aggregate equilibrium pay-off of all players is maximized (subject to the incentive constraints) if and only if aggregate experimentation is maximized (subject to the incentive constraints). Furthermore, we can characterize how maximum equilibrium experimentation is achieved. Interestingly, it requires alternation between pure-strategy and mixed-strategy equilibria of the stage game: for very low values of the prior, exactly one player experiments; as common beliefs rise, two players mix; then exactly two players experiment; then three players mix; and so on. Since players' best responses can be calculated explicitly, it is straightforward to characterize the cut-offs for the prior beliefs where the equilibrium switches from nobody choosing the risky action to experimentation by one player, and then from experimentation by one player to randomization by two players, etc.

The maximal equilibria can be understood intuitively as follows. When the opportunity cost of experimentation is high, even a small amount of experimentation by the other players will deter the remaining player from experimenting. The best way of maximizing total experimentation is therefore to have the other players refrain from experimentation altogether, leaving the remaining player to undertake a full unit of experimentation. As the opportunity cost of experimentation falls, the advantage of having two players contribute to experimentation outweighs the disadvantage of the disincentive to one player that results from experimentation by the other. The best way of maximizing total experimentation is therefore to have two players mix. As the opportunity cost of experimentation falls further, it becomes feasible to persuade one player to experiment even when a second player is already experimenting for sure. The best way to maximize experimentation is therefore to have two players experiment for sure. And so on. In particular, when the opportunity cost of experimentation is high, we see a single pioneer shouldering the burden of experimentation. As the opportunity cost of experimentation falls, two players share the burden. And so on.

The main reason why the undiscounted case is easier to analyse is that, in this case, players' best responses can be calculated without first solving for their value functions. By contrast, in the discounted case, players' best responses do depend on their value functions. Hence, since closed-form solutions for the value functions cannot be obtained in the discounted case, it is not possible to provide a complete characterization of equilibrium experimentation.

Besides providing a much more detailed characterization of equilibria of the strategic experimentation game, we believe that our paper also outlines a simple methodology for characterizing equilibria of complex stochastic differential games in the special case where there is no discounting. This methodology had already been put to use in Harris (1993); it greatly facilitated our analysis of the strategic-experimentation game in Bolton and Harris (1993 and 1996); and it has subsequently been put to use in Bergemann and Valimaki (1997) and Keller and Rady (1999).

The paper is organized as follows. Section 2 describes the basic model and introduces the background signal. Section 3 defines players' objectives and shows how players' preferences converge to a well-defined limit as the discount rate converges to zero. Section 4 characterizes the dynamics of beliefs for any profile of experimentation strategies. Section 5 defines best responses and equilibrium. Section 6 characterizes best responses. Section 7 characterizes the set of equilibria. Finally, Section 8 offers some concluding remarks.

2. The Model

There are N identical players. At the outset of the game, they all believe that $\mu = l$ with probability $1 - p_0$ and $\mu = h$ with probability p_0. At time t, the players simultaneously and independently choose between two actions, action 0 (the safe action) and action 1 (the risky action). If player i chooses action a_i, then her pay-off is

$$d\pi_i(t) = \begin{Bmatrix} sdt + \sigma dZ_i(t) & \text{if } a_i = 0 \\ \mu dt + \sigma dZ_i(t) & \text{if } a_i = 1 \end{Bmatrix}.$$

All players then observe all the actions chosen and all the resulting pay-offs. They also observe a background signal

$$d\pi_0(t) = \sqrt{(x_0)}\mu dt + \sigma dZ_0(t).$$

Here: s is fixed and known; $l < s < h$; $x_0 > 0$ is the quality of the background signal; and the $dZ_i(t)$ are independently and normally distributed with mean 0 and variance dt for $0 \leq i \leq N$.

3. The Objective

The objective of player i can be stated informally as follows: maximize the expectation of the undiscounted integral of the pay-offs $d\pi_i(t)$. In order to arrive at a formal statement that embodies this objective, one can proceed in one of two ways.

The first approach is to consider the limit of the expectation of the discounted integral of the pay-offs $d\pi_i(t)$ as the discount rate goes to 0. Suppose that, at time t: the players believe that $\mu = h$ with probability $p(t)$; and player i chooses the action $x_i(t) \in \{0, 1\}$. Define the full-information pay-off $\bar{u}$ by the formula

$$\bar{u}(q) = (1 - q)s + qh,$$

and define the expected risky pay-off m by the formula

$$m(q) = (1 - q)l + qh.$$

Then:

Proposition 1. *We have*

$$\lim_{r \to 0+} \frac{1}{r}\left(\mathrm{E}\left[\int_0^\infty re^{-rt} d\pi_i(t)\right] - \bar{u}(p_0)\right)$$

$$= \mathrm{E}\left[\int_0^\infty ((1 - x_i(t))s + x_i(t)m(p(t)) - \bar{u}(p(t)))dt\right].$$

This proposition can be understood as follows. The background information provided by the signal π_0 ensures that players eventually learn the value of μ. Hence, for any reasonable strategy of player i,

$$\lim_{r\to 0+} \mathrm{E}\left[\int_0^\infty re^{-rt}d\pi_i(t)\right] = \bar{u}(p_0).$$

For example, this is true of any Markov strategy that: (i) assigns probability 0 to the risky action when $p = 0$; (ii) assigns probability 1 to the risky action when $p = 1$; and (iii) is continuous at both $p = 0$ and $p = 1$.[1] Hence, for r near to 0, player i's preferences over her reasonable strategies are well represented by

$$\lim_{r\to 0+} \frac{1}{r}\left(\mathrm{E}\left[\int_0^\infty re^{-rt}d\pi_i(t)\right] - \bar{u}(p_0)\right)$$

wherever this limit exists. Finally, this limit exists for all strategies of player i, reasonable or not.

Proof. Let us denote the information available at time t by $\mathcal{F}_t$. We have

$$\begin{aligned}&\mathrm{E}[x_i(t)\mu \mid \mathcal{F}_t] = \mathrm{E}[x_i(t)\mid \mathcal{F}_t]\mathrm{E}[\mu \mid \mathcal{F}_t]\\ &= \mathrm{E}[x_i(t)\mid \mathcal{F}_t]m(p(t)) = \mathrm{E}[x_i(t)m(p(t))\mid \mathcal{F}_t]\end{aligned}$$

(because $x_i(t)$ is conditionally independent of μ given $\mathcal{F}_t$) and

$$\mathrm{E}[\sigma dZ_i(t)\mid \mathcal{F}_t] = 0.$$

Hence

$$\begin{aligned}&\mathrm{E}[d\pi_i(t)\mid \mathcal{F}_t]\\ &= \mathrm{E}[(1 - x_i(t))(sdt + \sigma dZ_i(t)) + x_i(t)(\mu dt + \sigma dZ_i(t))\mid \mathcal{F}_t]\\ &= \mathrm{E}[(1 - x_i(t))sdt + x_i(t)\mu dt + \sigma dZ_i(t)\mid \mathcal{F}_t]\\ &= \mathrm{E}[(1 - x_i(t))sdt + x_i(t)m(p(t))dt\mid \mathcal{F}_t].\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{E}\left[\int_0^\infty re^{-rt}d\pi_i(t)\right] &= \mathrm{E}\left[\int_0^\infty re^{-rt}\mathrm{E}[d\pi_i(t)\mid \mathcal{F}_t]\right]\\ &= \mathrm{E}\left[\int_0^\infty re^{-rt}\mathrm{E}[(1 - x_i(t))sdt + x_i(t)m(p(t))dt\mid \mathcal{F}_t]\right]\\ &= \mathrm{E}\left[\int_0^\infty re^{-rt}((1 - x_i(t))sdt + x_i(t)m(p(t))dt)\right]\end{aligned}$$

[1] For the formal definition of a Markov strategy, see Definition 5 below.

$$= \mathrm{E}\left[\int_0^\infty re^{-rt}((1 - x_i(t))s + x_i(t)m(p(t)))dt\right].$$

We also have

$$\bar{u}(p_0) = \mathrm{E}[\bar{u}(p(t))]$$

(because $p(t)$ follows a martingale and $\bar{u}$ is linear). Hence

$$\bar{u}(p_0) = \mathrm{E}\left[\int_0^\infty re^{-rt}\bar{u}(p_0)dt\right] = \mathrm{E}\left[\int_0^\infty re^{-rt}\bar{u}(p(t))dt\right].$$

Overall, then,

$$\frac{1}{r}\left(\mathrm{E}\left[\int_0^\infty re^{-rt}d\pi_i(t)\right] - \bar{u}(p_0)\right)$$

$$= \mathrm{E}\left[\int_0^\infty e^{-rt}((1 - x_i(t))s + x_i(t)m(p(t)) - \bar{u}(p(t)))dt\right].$$

But

$$(1 - x_i(t))s + x_i(t)m(p(t)) - \bar{u}(p(t)) \leqslant 0.$$

Letting $r \to 0$ and applying the monotone convergence theorem, we therefore obtain the required conclusion. ■

The second approach is to consider the limit of the expectation of the mean of the pay-offs $d\pi_i(t)$ over the interval $[0, T]$ as T goes to $+\infty$. On this approach:

Proposition 2. *We have*

$$\lim_{T\to+\infty} T\left(\mathrm{E}\left[\frac{1}{T}\int_0^T d\pi_i(t)\right] - \bar{u}(p_0)\right)$$

$$= \mathrm{E}\left[\int_0^\infty ((1 - x_i(t))s + x_i(t)m(p(t)) - \bar{u}(p(t)))dt\right].$$

This proposition can be understood in much the same way as Proposition 1. For any reasonable strategy of player i,

$$\lim_{T\to+\infty} \mathrm{E}\left[\frac{1}{T}\int_0^T d\pi_i(t)\right] = \bar{u}(p_0).$$

Hence, for T near to $+\infty$, player i's preferences over her reasonable strategies are well represented by

$$\lim_{T\to+\infty} T\left(\mathrm{E}\left[\frac{1}{T}\int_0^T d\pi_i(t)\right] - \bar{u}(p_0)\right),$$

wherever this limit exists. Finally, this limit exists for all strategies of player i, reasonable or not. The proof, which is analogous to that of Proposition 1, is omitted.

Propositions 1 and 2 lead us to the following definition:

Definition 3. *Player i's objective is to maximize*

$$\mathrm{E}\left[\int_0^\infty ((1 - x_i(t))s + x_i(t)m(p(t)) - \bar{u}(p(t)))dt\right].$$

4. The Dynamics of Beliefs

Suppose that, at the outset of period t, the players believe that $\mu = h$ with probability $p(t)$. Let $x_0(t) = x_0$, and suppose that the players $i \in \{1, \ldots, N\}$ choose the actions $x_i(t) \in \{0, 1\}$. Suppose further that, at the conclusion of period t, the players believe that $\mu = h$ with probability $p(t + dt)$. Let

$$dp(t) = p(t + dt) - p(t)$$

denote the change in beliefs concerning μ; and define the information function Φ by the formula

$$\Phi(q) = \left(\frac{q(1 - q)(h - l)}{\sigma}\right)^2.$$

Then:

Proposition 4. *We have*

$$dp(t) = p(t)(1 - p(t))\left(\frac{h - l}{\sigma}\right)\sum_{i=0}^{N}\sqrt{(x_i(t))}d\tilde{Z}_i(t),$$

where

$$d\tilde{Z}_i(t) = \frac{1}{\sigma}(\sqrt{(x_i(t))}(\mu - m(p(t)))dt + \sigma dZ_i(t)).$$

Moreover, conditional on the information available at time t, the $d\tilde{Z}_i(t)$ *are independently and identically distributed with mean* 0 *and variance dt.*

In particular, conditional on the information available at time t, $dp(t)$ has mean 0 and variance $(\sum_{i=0}^{N} x_i(t))\Phi(p(t))dt$.

Proof. See Lemma 1 of Bolton and Harris (1999). ■

5. Definition of Equilibrium

Since the only pay-off relevant variable is the probability that $\mu = h$, and since a player's mixed action can be identified with the probability with which she chooses action 1, both the state space and the mixed-action set for our model can be taken to be the unit interval [0, 1]. Hence:

Definition 5. *A* **Markov strategy** *for player i is a Borel measurable function* $\xi_i : [0, 1] \to [0, 1]$.

It can be shown that, for all Markov-strategy profiles and all initial beliefs, there is a unique solution to the dynamics. More precisely, put $I = \{1, 2, \dots, N\}$, put $\xi_0(q) = x_0$ for all $q \in [0, 1]$, let $C([0, +\infty), [0, 1])$ denote the space of continuous functions from the time line $[0, +\infty)$ to the state space [0, 1], let $\mathcal{P}(C([0, +\infty), [0, 1]))$ denote the space of probability measures on $C([0, +\infty), [0, 1])$, and let p denote the identity mapping on $C([0, +\infty), [0, 1])$. Then:

Proposition 6. *For all Markov-strategy profiles* $\xi = \times_{i\in I}\xi_i$ *and all initial beliefs* $p_0 \in [0, 1]$, *there is a unique* $\lambda(p_0) \in \mathcal{P}(C([0, +\infty), [0, 1]))$ *such that, with* $\lambda(p_0)$*-probability one:*

(i) $p(0) = p_0$; *and*
(ii) for all $t \in [0, +\infty)$, $dp(t)$ *has mean* 0 *and variance* $(\sum_{i=0}^{N}\xi_i(p(t)))\Phi(p(t))dt$.

Proof. This follows at once from the results of Engelbert and Schmidt (1985), as described in Section 5.5 of Karatzas and Shreve (1988). ■

Proposition 6 implies that, for all Markov-strategy profiles ξ and all initial beliefs p_0, player i's expected pay-off $g_i(\xi, p_0)$ is well defined. Hence:

Definition 7. *The Markov strategy* ξ_i *is a* **perfect best response** *to the Markov-strategy profile* $\xi_{-i} = \times_{j\in I\setminus\{i\}}\xi_j$ *iff, for all* $p_0 \in [0, 1]$, ξ_i *is a best response to* ξ_{-i} *in the game in which player i has pay-off function* $g_i(\cdot, p_0)$.

Definition 8. *The Markov-strategy profile* $\xi = \times_{i\in I}\xi_i$ *is an* **equilibrium** *iff, for all* $i \in \{1, 2, \dots, N\}$, ξ_i *is a perfect best response to* $\xi_{-i} = \times_{j\in I\setminus\{i\}}\xi_j$.

6. Characterization of Best Responses

Suppose that the players $j \in I\setminus\{i\}$ employ the Markov strategies ξ_j, and let $\Xi_{-i} = \sum_{j\in(\{0\}\cup I)\setminus\{i\}}\xi_j$. Then:

Proposition 9. *Player i's value function* $u_i : [0, 1] \to (-\infty, 0]$ *satisfies*

$$0 = \max_{\alpha_i\in[0,1]}\left\{(1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{u_i''(p)}{2}\right\} \tag{1}$$

for all $p \in (0, 1)$ *and*

$$u_i(0) = 0,\ u_i(1) = 0. \tag{2}$$

Moreover the Markov strategy ξ_i *is a perfect best response for player i iff*

$$\xi_i(p) \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \left\{ (1 - \alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{u_i''(p)}{2} \right\} \tag{3}$$

for all $p \in (0, 1)$ *and*

$$\xi_i(0) = 0,\ \xi_i(1) = 1. \tag{4}$$

Proof. Let $H_i(p, \alpha_i, \xi_{-i}, c_i)$ denote the expectation of player i's current pay-off when the current belief is $p \in [0, 1]$, player i chooses the risky action with probability $\alpha_i \in [0, 1]$, the profile of Markov strategies employed by the other players is ξ_{-i} and player i's continuation pay-offs are given by the function $c_i : [0, 1] \to (-\infty, 0]$. Then player i's value function u_i must satisfy the Bellman equation

$$u_i(p) = \max_{\alpha_i \in [0,1]} \{H_i(\alpha_i, p, u_i, \xi_{-i})\} \tag{5}$$

for all $p \in [0, 1]$. Moreover the strategy ξ_i is a perfect best response for player i iff

$$\xi_i(p) \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \{H_i(\alpha_i, p, u_i, \xi_{-i})\} \tag{6}$$

for all $p \in [0, 1]$.

Now suppose that the realized actions are $\{x_j \mid 1 \leqslant j \leqslant N\}$, and put $X_i = \sum_{j \in (\{0\} \cup I) \setminus \{i\}} x_j$. Then player i's current pay-off is

$$((1 - x_i)s + x_i m(p) - \bar{u}(p))dt$$

(by Definition 3) and her continuation pay-off is

$$c_i(p + dp) = c_i(p) + c_i'(p)dp + \frac{1}{2}c_i''(p)dp^2$$

(by Itô's Lemma), where dp has mean 0 and variance $(x_i + X_i)\Phi(p)dt$ (by Proposition 4). Hence the expectation of her current pay-off conditional on $\{x_j \mid 1 \leqslant j \leqslant N\}$ is

$$c_i(p) + \left((1 - x_i)s + x_i m(p) - \bar{u}(p) + (x_i + X_i)\Phi(p)\frac{c_i''(p)}{2} \right) dt.$$

Hence

$$\begin{aligned} &H_i(\alpha_i, p, c_i, \xi_{-i}) = c_i(p) \\ &+ \left((1 - \alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{c_i''(p)}{2} \right) dt \end{aligned} \tag{7}$$

(on taking expectations over the x_j).

Next, combining (5) and (7), we obtain

$$u_i(p)$$

$$= \max_{\alpha_i \in [0,1]} \left\{ u_i(p) + \left((1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{u_i''(p)}{2} \right) dt \right\}$$

$$= u_i(p) + \left(\max_{\alpha_i \in [0,1]} \left\{ (1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{u_i''(p)}{2} \right\} \right) dt.$$

Hence, subtracting $u_i(p)$ from both sides and dividing both sides by dt,

$$0 = \max_{\alpha_i \in [0,1]} \left\{ (1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{u_i''(p)}{2} \right\}$$

for all $p \in [0, 1]$. Similarly, combining (6) and (7), we obtain

$$\xi_i(p) \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \left\{ (1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p) + (\alpha_i + \Xi_{-i}(p))\Phi(p)\frac{u_i''(p)}{2} \right\} \quad (8)$$

for all $p \in [0, 1]$.

Finally, using (8) and noting that $\Phi(0) = \Phi(1) = 0$, we obtain

$$\xi_i(0) \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \{(1-\alpha_i)s + \alpha_i m(0) - \bar{u}(0)\} = \operatorname*{argmax}_{\alpha_i \in [0,1]} \{-\alpha_i(s-l)\}$$

and

$$\xi_i(1) \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \{(1-\alpha_i)s + \alpha_i m(1) - \bar{u}(1)\} = \operatorname*{argmax}_{\alpha_i \in [0,1]} \{-(1-\alpha_i)(h-s)\}.$$

Hence $\xi_i(0) = 0$ and $\xi_i(1) = 1$. Hence the flow pay-offs in states 0 and 1 are

$$(1 - \xi_i(0))s + \xi_i(0)m(0) - \bar{u}(0) = -\xi_i(0)(s - l) = 0$$

and

$$(1-\xi_i(1))s + \xi_i(1)m(1) - \bar{u}(1) = -(1 - \xi_i(1))(h - s) = 0.$$

Hence $u_i(0) = 0$ and $u_i(1) = 0$. ■

The following lemma simplifies both the problem of finding player i's value function and the problem of finding player i's perfect best responses.

Lemma 10. *Equation (1) holds iff*

$$0 = \max_{\alpha_i \in [0,1]} \left\{ \frac{(1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p)}{\alpha_i + \Xi_{-i}(p)(p)} \right\} + \Phi(p)\frac{u_i''(p)}{2}. \quad (9)$$

Moreover, if either (1) or (9) holds, then (3) holds iff

$$\xi_i(p) \in \underset{\alpha_i \in [0,1]}{\operatorname{argmax}} \left\{ \frac{(1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p)}{\alpha_i + \Xi_{-i}(p)(p)} \right\}. \tag{10}$$

Proof. This follows at once from the fact that $0 < x_0 \leqslant \alpha_i + \Xi_{-i} \leqslant x_0 + N < +\infty$. ■

Proposition 11. *There is a unique u_i satisfying equations (1 and 2).*

Proof. In view of Lemma 10, we need only show that there is a unique u_i such that

$$0 = \max_{\alpha_i \in [0,1]} \left\{ \frac{(1-\alpha_i)s + \alpha_i m(p) - \bar{u}(p)}{\alpha_i + \Xi_{-i}(p)} \right\} + \Phi(p)\frac{u_i''(p)}{2} \tag{11}$$

for all $p \in (0, 1)$ and

$$u_i(0) = 0, \ u_i(1) = 0. \tag{12}$$

Note first that, for any bounded $f\colon [0, 1] \to \mathbb{R}$,

$$0 = f(p) + \Phi\,(p)\frac{v''(p)}{2}$$

for all $p \in (0, 1)$ iff

$$v(p) = \int_0^1 G(p, q)f(q)dq + (1 - p)v(0) + pv(1)$$

for all $p \in [0, 1]$, where

$$G(p, q) = \begin{cases} \dfrac{2\sigma^2 p}{(h-l)^2 q^2 (1-q)} & \text{if } p \in [0, q] \\[2ex] \dfrac{2\sigma^2 (1-p)}{(h-l)^2 q (1-q)^2} & \text{if } p \in [q, 1] \end{cases}$$

(by standard considerations in the theory of differential equations). Hence equations (11 and 12) hold iff

$$u_i(p) = \int_0^1 G(p, q)f(q)dq,$$

where

$$f(q) = \max_{\alpha_i \in [0,1]} \left\{ \frac{(1-\alpha_i)s + \alpha_i m(q) - \bar{u}(q)}{\alpha_i + \Xi_{-i}(q)} \right\}.$$

Noting that neither G nor f involve u_i, we conclude that there is a unique solution of equations (11 and 12). ■

Define the break-even probability b and the incentive to experiment β by the formulae

$$b = \frac{s - l}{h - l}$$

and

$$\beta(p) = \begin{Bmatrix} \dfrac{\bar{u}(p) - s}{s - m(p)} & \text{if } p \in [0, b) \\ +\infty & \text{if } p \in [b, 1] \end{Bmatrix}.$$

Then:

Proposition 12. *The Markov strategy ξ_i is a perfect best response for player i iff*

$$\xi_i(p) \begin{Bmatrix} = 0 & \text{if } \beta(p) < \Xi_{-i}(p) \\ \in [0, 1] & \text{if } \beta(p) = \Xi_{-i}(p) \\ = 1 & \text{if } \beta(p) > \Xi_{-i}(p) \end{Bmatrix}$$

for all $p \in [0, 1]$.

In other words, player i should experiment iff the incentive to experiment exceeds the total experimentation by the other players (including an allowance for the amount of experimentation effectively contributed by the background signal).

Proof. Elementary manipulation shows that

$$\frac{(1 - \alpha_i)s + \alpha_i m(p) - \bar{u}(p)}{\alpha_i + \Xi_{-i}(p)} = -(s - m(p)) - \frac{(\bar{u}(p) - s) - \Xi_{-i}(p)(s - m(p))}{\alpha_i + \Xi_{-i}(p)}.$$

Moreover: if $p < b$ then $s - m(p) > 0$, and so

$$(\bar{u}(p) - s) - \Xi_{-i}(p)(s - m(p)) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 0 \text{ iff } \frac{\bar{u}(p) - s}{s - m(p)} \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \Xi_{-i}(p);$$

and if $p \geqslant b$ then $s - m(p) \leqslant 0$, and so

$$(\bar{u}(p) - s) - \Xi_{-i}(p)(s - m(p)) \geqslant \bar{u}(p) - s > 0. \;■$$

7. Characterization of Equilibrium

Proposition 12 allows us to give a complete classification of the equilibria of our game. For all mixed-action profiles $\alpha = \times_{i=1}^{N}\alpha_i$, let k_0 be the number of players who play the safe action with probability one, let k_1 be the number of players who play the risky action with probability one, and let k_M be the number of players who play a strictly mixed action. For all $\gamma \in [0, +\infty]$, let $E(\gamma)$ be the set of mixed-action profiles α such that one of the following conditions is satisfied:

(i) $k_M = 0$, $k_1 = 0$ and $\gamma \leqslant x_0$;
(ii) $k_M = 0$, $1 \leqslant k_1 \leqslant N - 1$ and $x_0 + k_1 - 1 \leqslant \gamma \leqslant x_0 + k_1$;
(iii) $k_M = 0$, $k_1 = N$ and $\gamma \geqslant x_0 + N - 1$;
(iv) $k_M = 1$, $\gamma = x_0 + k_1$ and the player who mixes chooses the risky action with any probability in (0, 1);
(v) $k_M \geqslant 2$, $x_0 + k_1 < \gamma < x_0 + k_1 + k_M - 1$ and the players who mix all choose the risky action with probability $(y - x_0 - k_1)/(k_M - 1)$.

In other words, either the incentive to experiment is so low that having no players experiment is incentive-compatible; or it is sufficiently high that at least k_1 players can be persuaded to experiment, but sufficiently low that at most k_1 players can be persuaded to experiment; or it is so high that all players can be persuaded to experiment; or it takes on a knife-edge value at which a $(k_1 + 1)^{\text{th}}$ player is indifferent between experimenting and not experimenting; or it lies in a range in which k_M players can be persuaded to mix when k_1 players experiment, provided that the probability with which the mixing-players experiment is chosen appropriately. Then:

Proposition 13. *The Markov-strategy profile* $\xi = \times_{i=1}^{N}\xi_i$ *is an equilibrium iff, for all* $p \in [0, 1]$, $\xi(p) \in E(\beta(p))$.

In other words, ξ is an equilibrium iff $\xi(p)$ is a Nash equilibrium of the stage game for all $p \in [0, 1]$.

Proof. Fix $p \in [0, 1]$, put $\alpha = \xi(p)$ and put $\gamma = \beta(p)$. Note first that, if $k_M = 0$ and $k_1 \geqslant 1$, then playing risky is incentive-compatible for those players who play risky iff $\gamma \geqslant x_0 + k_1 - 1$. Similarly, if $k_M = 0$ and $k_1 \leqslant N - 1$, then playing safe is incentive-compatible for those players who play safe iff $\gamma \leqslant x_0 + k_1$. Secondly, if $k_M = 1$ then mixing is incentive-compatible for the player who mixes iff $\gamma = x_0 + k_1$. Moreover: playing safe is less attractive for the players who play risky than it is for the player who mixes; and playing risky is less attractive for the players who play safe than it is for the player who mixes. Thirdly, if $k_M \geqslant 2$, then mixing is incentive-compatible for player i iff $\gamma = \Xi_{-i} = (\sum_{j=0}^{N}\xi_j) - \xi_i$. Hence mixing is incentive-compatible for those players who mix iff they all mix with the same probability $(y - x_0 - k_1)/(k_M - 1)$. Moreover: $(y - x_0 - k_1)/(k_M - 1) \in (0, 1)$ iff $x_0 + k_1 < \gamma < x_0 + k_1 + k_M - 1$; playing safe is less attractive for the players who play risky than it is for the players who mix; and playing risky is less attractive for the players who play safe than it is for the players who mix. ■

Proposition 13 in turn allows us to identify those equilibria that maximize aggregate pay-offs. Let $g(\xi, p_0)$ denote the aggregate pay-off when the Markov-strategy profile ξ is employed and the initial belief is p_0, let $\mathcal{E}$ denote the set of equilibria, let

$$\bar{\mathcal{E}}(p_0) = \underset{\xi \in \mathcal{E}}{\operatorname{argmax}} \{g(\xi, p_0)\},$$

and define $\bar{E} : [0, +\infty] \to [0, N]$ by the formula

$$\bar{E}(\gamma) = \underset{\alpha \in E(\gamma)}{\operatorname{argmax}} \left\{ \sum_{i \in I} \alpha_i \right\}.$$

Then:

Proposition 14. *For all $p_0 \in [0, 1]$, $\xi \in \bar{\mathcal{E}}(p_0)$ iff $\xi(p) \in \bar{E}(\beta(p))$ for all $p \in [0, 1]$.*

In other words, for any given initial belief, an equilibrium maximizes the aggregate pay-off iff it maximizes total experimentation in every state. In particular, if an equilibrium maximizes the aggregate pay-off for some initial belief, then it maximizes the aggregate pay-off for all initial beliefs.

Proof. Suppose that ξ is an equilibrium. Then Proposition 9 and Lemma 10 imply that

$$0 = \frac{(1 - \xi_i(p))s + \xi_i(p)m(p) - \bar{u}(p)}{\xi_i(p) + \Xi_{-i}(p)} + \Phi(p)\frac{u_i''(p)}{2} \tag{13}$$

for $p \in (0, 1)$ and

$$u_i(0) = u_i(1) = 0. \tag{14}$$

Put $v = \sum_{i \in I} u_i$. Then, summing (13) and (14) over i, we obtain

$$0 = \frac{-N(\bar{u}(p) - s) - (s - m(p))\sum_{i \in I}\xi_i(p)}{x_0 + \sum_{i \in I}\xi_i(p)} + \Phi(p)\frac{v''(p)}{2} \tag{15}$$

for $p \in (0, 1)$ and

$$v(0) = v(1) = 0. \tag{16}$$

Moreover, rearranging (15), we obtain

$$0 = -(s - m(p)) - \frac{N(\bar{u}(p) - s) - x_0(s - m(p))}{x_0 + \sum_{i \in I}\xi_i(p)} + \Phi(p)\frac{v''(p)}{2}.$$

As in the proof of Proposition 11, it follows that (15 and 16) hold iff

$$v(p) = \int_0^1 G(p, q) f(q, \xi(q))dq + (1 - p)v(0) + pv(1)$$

for $p \in [0, 1]$, where

$$G(p, q) = \left\{ \begin{array}{ll} \dfrac{2\sigma^2 p}{(h-l)^2 q^2 (1-q)} & \text{if } p \in [0, q] \\ \dfrac{2\sigma^2 (1-p)}{(h-l)^2 q (1-q)^2} & \text{if } p \in [q, 1] \end{array} \right\}$$

and

$$f(q, \alpha) = -(s - m(q)) - \frac{N(\bar{u}(q) - s) - x_0(s - m(q))}{x_0 + \sum_{i \in I} \alpha_i}.$$

It follows at once that $\xi \in \bar{E}(p_0)$ iff

$$\xi(q) \in \underset{\alpha \in E(q)}{\operatorname{argmax}} \{f(q, \alpha)\}$$

for $q \in [0, 1]$. Finally, $\xi_i(q) = 0$ whenever $\beta(q) < x_0$,

$$N(\bar{u}(q) - s) - x_0(s - m(q)) > 0$$

whenever $\beta(q) \in [x_0, x_0 + N - 1]$, and $\xi_i(q) = 1$ whenever $\beta(q) > x_0 + N - 1$. Hence

$$\underset{\alpha \in E(q)}{\operatorname{argmax}} \{f(q, \alpha)\} = \bar{E}(\beta(q)).$$

This completes the proof. ■

Figure 4.1 depicts the total-experimentation correspondence $\hat{E}$ given by the formula $\hat{E}(\gamma) = \{\sum_{i \in I} \alpha_i \mid \alpha \in E(\gamma)\}$ in the case $N = 3$ and $x_0 = 2$. It suggests that, in order to maximize total experimentation: there should initially be a single pioneer who

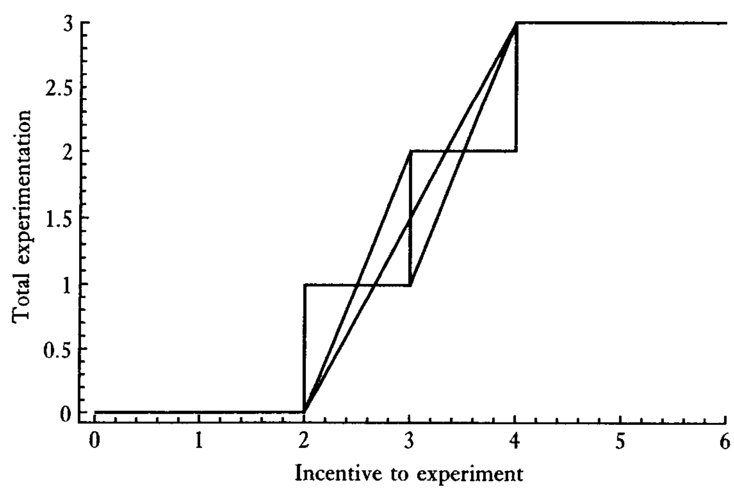

FIG. 4.1. *The total-experimentation correspondence*

experiments alone; then two pioneers should share the burden of experimentation; then three pioneers should share the burden of experimentation; and so on.

The following proposition gives an explicit characterization of $\bar{E}(\gamma)$.

Proposition 15. *For all* $l \in \{1, 2, \ldots, N-1\}$:

(i) *for all* $\gamma \in [x_0 + l - 1,\ x_0 + l - 1 + \frac{1}{l+1})$, $\alpha \in \bar{E}(\gamma)$ *iff* $k_M = 0$ *and* $k_1 = l$;

(ii) *for all* $\gamma \in (x_0 + l - 1 + \frac{1}{l+1},\ x_0 + l)$, $\alpha \in \bar{E}(\gamma)$ *iff* $k_M = l + 1$ *and* $k_1 = 0$;

and, in the knife-edge case in which $\gamma = x_0 + l - 1 + \frac{1}{l+1}$, $\alpha \in \bar{E}(\gamma)$ *iff either* $k_M = 0$ *and* $k_1 = l$, *or* $k_M = l + 1$ *and* $k_1 = 0$.

In particular: for $\gamma \in [x_0, x_0 + \frac{1}{2})$, exactly one player plays risky; for $\gamma \in (x_0 + \frac{1}{2}, x_0 + 1)$, exactly two players randomize; for $\gamma \in [x_0 + 1, x_0 + 1 + \frac{1}{3})$, exactly two players play risky; for $\gamma \in (x_0 + 1 + \frac{1}{3}, x_0 + 2)$, exactly three players randomize; and so on up to $l = N - 1$.

Proof. Note first that any $\alpha \in E(\gamma)$ in which both $k_1 > 0$ and $k_M > 0$ is dominated by the corresponding equilibrium $\hat{\alpha} \in E(\gamma)$ in which $\hat{k}_1 = 0$ and $\hat{k}_M = k_1 + k_M$. Hence we can confine attention to equilibria in which either $k_1 = 0$ or $k_M = 0$. Secondly, any $\alpha \in E(\gamma)$ in which $k_1 = 0$ and $k_M \geqslant 3$ is dominated by the corresponding equilibrium $\hat{\alpha} \in E(\gamma)$ in which $\hat{k}_1 = 0$ and $\hat{k}_M = k_M - 1$ if $\gamma \in (x_0, x_0 + k_M - 2)$. Hence equilibria in which $k_1 = 0$ and $k_M \geqslant 2$ are only relevant when $\gamma \in (x_0 + k_M - 2, x_0 + k_M - 1)$. Thirdly, any $\alpha \in E(\gamma)$ in which $k_1 \leqslant N - 1$ and $k_M = 0$ is dominated by the corresponding equilibrium $\hat{\alpha} \in E(\gamma)$ in which $\hat{k}_1 = k_1 + 1$ and $\hat{k}_M = 0$ if $\gamma = x_0 + k_1$. Hence equilibria in which $k_1 \leqslant N - 1$ and $k_M = 0$ are only relevant when $\gamma \in [x_0 + k_1 - 1, x_0 + k_1)$. Finally, for all $l \in \{1, 2, \ldots, N-1\}$ and all $\gamma \in [x_0 + l - 1, x_0 + l)$, the equilibrium in which $k_1 = l$ and $k_M = 0$ dominates the equilibrium in which $k_1 = 0$ and $k_M = l + 1$ if $\gamma \in [x_0 + l - 1, x_0 + l - 1 + \frac{1}{l+1}]$, and vice versa if $\gamma \in (x_0 + l - 1 + \frac{1}{l+1}, x_0 + l)$. ■

8. Conclusion

As we have seen, the characterization of perfect best responses becomes very simple in the limit case where there is no discounting. As a result, it is possible to provide a complete classification of equilibria for this limiting case. The analysis of this paper therefore provides an illustration of the important simplifications that may be obtainable in the case of no discounting for the analysis of stochastic differential games. We believe that this methodology can be applied to problems other than experimentation and learning, and may provide a useful key to solve problems that have previously been thought to be intractable.

References

Bergemann, D., and Valimaki, J. (1997) 'Market Diffusion with Two-sided Learning', *RAND Journal of Economics* 28: 773–95.

—— and Valimaki, J. (2000) 'Experimentation in Markets', *Review of Economic Studies* (forthcoming).

Berry, D. A., and Fristedt, B. (1985) *Bandit Problems*. Chapman & Hall: New York.

Bolton, P., and Harris, C. (1993) 'Strategic Experimentation', STICERD Theoretical Economics Discussion Paper TE/93/261, London School of Economics, January.

—— —— (1996) 'Strategic Experimentation: A Revision', Discussion Paper 9627, CentER Tilburg University.

—— —— (1999) 'Strategic Experimentation', *Econometrica* 67: 349–74.

Engelbert, H. J., and Schmidt, W. (1985) 'On Solutions of Stochastic Differential Equations without Drift', *Zeitschrift für Wahrscheinlichskeitstheorie und verwandte Gebiete* 68: 287–317.

Harris, C. (1993), 'Generalized Solutions of Stochastic Differential Games in One Dimension', Industry Studies Program Discussion Paper 44, Boston University, December.

Karatzas, I. (1984) 'Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes', *Annals of Probability* 12: 173–92.

—— and Shreve, S. E. (1988) *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

Keller, G., and Rady, S. (1999) 'Optimal Experimentation in a Changing Environment', *Review of Economic Studies* 66: 475–509.

Krylov, N. V. (1980) *Controlled Diffusion Processes*. Springer-Verlag, New York.

Mandelbaum, A. (1987) 'Continuous Multi-Armed Bandits and Multi-Parameter Processes', *Annals of Probability* 15: 1527–56.

Smith, L. (1991) 'Error Persistence and Experimental versus Observational Learning', Foerder Discussion Paper, Tel-Aviv University and MIT.